Elementary probability

DAVID STIRZAKER

Mathematical Institute, University of Oxford

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, United Kingdom
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1994

This book is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 1994

Reprinted 1995, 1997

Printed in the United Kingdom at the University Press, Cambridge

Typeset in 10/12½pt Times

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Stirzaker, David.
Elementary probability / David Stirzaker.
p. cm.
Includes index.
ISBN 0-521-42028-8 (hbk). – ISBN 0-521-42183-7 (pbk)
1. Probabilities. I. Title.
A273.S7534 1994
519.2–dc20 93-12917 CIP

ISBN 0 521 42028 8 hardback
ISBN 0 521 42183 7 paperback

Contents

Preface ix

1 **Probability** 1
- 1.1 Introduction 1
- 1.2 Events 1
- 1.3 Symmetry and probability 6
- 1.4 Properties of probability 8
- 1.5 Sequences of events 10
- 1.6 Remarks 11
- Worked examples and exercises 12
- Problems 19

2 **Conditional probability and independence** 22
- 2.1 Conditional probability 22
- 2.2 Independence 28
- 2.3 Recurrence and difference equations 31
- 2.4 Remarks 33
- Worked examples and exercises 33
- Problems 45

3 **Counting** 52
- 3.1 First principles 52
- 3.2 Permutations: ordered selection 53
- 3.3 Combinations: unordered selection 55
- 3.4 Inclusion–exclusion 56
- 3.5 Recurrence relations 57
- 3.6 Generating functions 58
- 3.7 Techniques 62
- Worked examples and exercises 64
- Problems 75

4 **Random variables: distribution and expectation** 81
- 4.1 Random variables 81
- 4.2 Distributions 82

4.3 Expectation 87
4.4 Conditional distributions 94
4.5 Sequences of distributions 97
4.6 Inequalities 98
Worked examples and exercises 100
Problems 115

5 Random vectors; independence and dependence 121
5.1 Joint distributions 121
5.2 Independence 125
5.3 Expectation 128
5.4 Sums and products of random variables 134
5.5 Dependence; conditional expectation 139
5.6 Simple random walk 145
5.7 Inequalities 151
5.8 The law of averages 152
5.9 Convergence 155
Worked examples and exercises 156
Problems 170

6 Generating functions and their applications 175
6.1 Introduction 175
6.2 Moments and the probability generating function 179
6.3 Sums of independent random variables 182
6.4 Moment generating functions 187
6.5 Joint generating functions 189
6.6 Sequences 193
6.7 Regeneration 196
6.8 Random walks 201
Worked examples and exercises 205
Problems 215

7 Continuous random variables 220
7.1 Density and distribution 220
7.2 Functions of random variables 230
7.3 Simulation of random variables 234
7.4 Expectation 235
7.5 Moment generating functions 239
7.6 Conditional distributions 242
7.7 Ageing and survival 244
7.8 Stochastic ordering 246
7.9 Random points 247
Worked examples and exercises 250
Problems 262

8 Jointly continuous random variables 265
8.1 Joint density and distribution 265
8.2 Change of variables 270
8.3 Independence 271
8.4 Sums, products and quotients 275
8.5 Expectation 278
8.6 Conditional density and expectation 282
8.7 Transformations; order statistics 287
8.8 The Poisson process 290
8.9 Two limit theorems 293
Worked examples and exercises 295
Problems 305

9 Markov chains 309
9.1 The Markov property 309
9.2 Transition probabilities 313
9.3 First passage times 318
9.4 Stationary distributions 324
9.5 The long run 330
9.6 Markov chains in general 336
9.7 Markov chains with continuous parameter 337
9.8 Forward equations; Poisson and birth processes 340
9.9 Forward equations; equilibrium 343
Worked examples and exercises 348
Problems 364

Appendix: Solutions and hints for selected exercises and problems 368

Further reading 401

Index 402

1

Probability

1.1 Introduction

The idea of probability underlies statements such as: 'Rain is likely today', 'This egg is probably infected with salmonella', and so on. These uses of probability can be expressed mathematically in various ways. Here, we start with the concept of an *experiment*, which is any procedure that has a well-defined set of outcomes; the set of outcomes is called the *sample space*. Subsets of the sample space are called *events*, and statements like those above typically assess how likely an event is to occur. The collection of events of interest is the *event space*, and *probability* is a function defined on each event in the event space. By convention a probability can be any number between zero and one, inclusive.

Of course the brief summary above calls for much explanation and elaboration. In the next few sections of this chapter we shall provide a few simple rules (or axioms) defining the properties of events and their probabilities. This choice of rules is guided by our experience of real events and their likelihoods, but our experience and intuition cannot *prove* that these rules are true or say what probability 'really' is. What we can say is that, starting with these rules, we can derive a theory that provides an elegant and accurate description of many random phenomena, ranging from the behaviour of queues in supermarkets to the behaviour of nuclear reactors.

1.2 Events

Suppose we are considering some experiment such as tossing a coin. To say that the experiment is well-defined means that we can list all the possible outcomes; in the case of a tossed coin the list reads: {head, tail}. For a general (unspecified) experiment any particular outcome is denoted by ω; the collection of all outcomes is called the *sample space* and is denoted by Ω.

Any specified collection of outcomes in Ω is called an *event*. Upper case letters such as A, B, C, are used to denote events; these may have suffices or other adornments such as A_i, $\bar{B}$, C^*, and so on. If the outcome of the experiment is ω, and $\omega \in A$, then A is said to *occur*. The set of outcomes not in A is called the complement of A, and is denoted by A^c.

In particular, the event which contains all possible outcomes is the *certain event*, and is denoted by Ω. Also, the event containing no outcomes is the *impossible event* and is denoted by ϕ. Obviously $\phi = \Omega^c$.

Example: Coins If a coin is tossed once, then $\Omega = \{\text{head, tail}\}$. In line with the notation above, we usually write

$$\Omega = \{H, T\}.$$

The event that the coin shows a head should strictly be denoted by $\{H\}$, but in common with most other writers we omit the braces in this case, and denote a head by H. Obviously $H^c = T$, and $T^c = H$.

Likewise, if a coin is tossed twice, then

$$\Omega = \{HH, HT, TH, TT\},$$

and so on. This experiment is performed even more often in probability textbooks than it is in real life. ●

Since events are sets we use the usual notation for combining them, thus:

$A \cap B$ denotes outcomes in both A and B, their *intersection*;
$A \cup B$ denotes outcomes in either A or B or both, their *union*;
$A \triangle B$ denotes outcomes in either A or B, but not both, their *symmetric difference*;
$A \backslash B$ denotes outcomes in A which are not in B, their *difference*;
$\bigcup_{j=1}^{\infty} A_j$ denotes outcomes which are in at least one of the countable collection $(A_j; j \geqslant 1)$, their *countable union*;
[Countable sets are in one–one correspondence with a subset of the positive integers.]
$A \subseteq B$ denotes that every outcome in A is also in B, this is *inclusion*;
$A = \{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}$ denotes that the event A consists of the outcomes $\omega_1, \dots, \omega_n$;
$A \times B$ denotes the product of A and B, that is the set of all ordered pairs (ω_a, ω_b) where $\omega_a \in A$ and $\omega_b \in B$.

These methods of combining events give rise to many equivalent ways of denoting an event. Some of the more useful identities for any events A and B are:

$$A \triangle B = (A \cap B^c) \cup (A^c \cap B) \tag{1}$$

$$A = (A \cap B) \cup (A \cap B^c) \tag{2}$$

$$A \backslash B = A \cap B^c \tag{3}$$

$$A^c = \Omega \backslash A \tag{4}$$

$$A \cap A^c = \phi \tag{5}$$

$$A \cup A^c = \Omega. \tag{6}$$

These are all easily verified by checking that every element of the left hand side is included in the right hand side, and vice versa. You should do this.

Such relationships are often very conveniently represented by simple diagrams.

We illustrate this by providing some basic examples in Figures 1.1 and 1.2. Similar relationships hold between combinations of three or more events and some of these are given in the problems at the end of this chapter.

When $A \cap B = \phi$ we say that A and B are disjoint (or mutually exclusive).

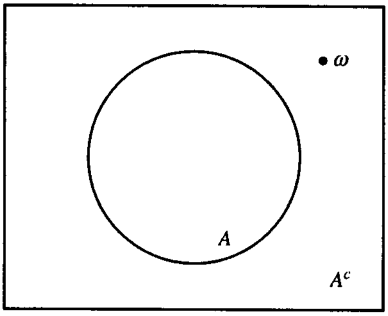

Figure 1.1 The interior of the rectangle represents the sample space Ω, and the interior of the circle represents an event A. The point ω represents an outcome in the event A^c. The diagram clearly illustrates the identities $A^c \cup A = \Omega$ and $\Omega \backslash A = A^c$.

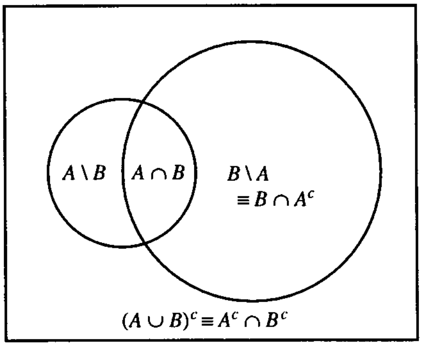

Figure 1.2 The interior of the smaller circle represents the event A; the interior of the larger circle represents the event B. The diagram illustrates a number of simple relationships; for example, the region common to both circles is $A \cap B \equiv (A^c \cup B^c)^c$. For another example, observe that $A \triangle B = A^c \triangle B^c$.

(7) **Example** A die* is rolled. The outcome is one of the integers from 1 to 6. We may denote these by $\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, or more directly by $\{1, 2, 3, 4, 5, 6\}$, as we choose. Define:

A the event that the outcome is even,
B the event that the outcome is odd,
C the event that the outcome is prime,
D the event that the outcome is perfect (a perfect number is the sum of its prime factors).

---

*Unless otherwise specified, a die is always a cube bearing the numbers 1, 2, 3, 4, 5 and 6 on its faces, one to each face.

Then the above notation compactly expresses obvious statements about these events. For example:

$$A \cap B = \phi \qquad A \cup B = \Omega$$

$$A \cap D = \{\omega_6\} \qquad C\backslash A = B\backslash\{\omega_1\}$$

and so on. ●

It is natural and often useful to consider the number of outcomes in an event A. This is denoted by $|A|$, and is called the *size* or *cardinality* of A.

It is straightforward to see, by counting the elements on each side, that size has the following properties.

If A and B are disjoint then

(8) $$|A \cup B| = |A| + |B|,$$

and more generally, for any A and B

(9) $$|A \cup B| + |A \cap B| = |A| + |B|.$$

If $A \subseteq B$ then

(10) $$|A| \leqslant |B|.$$

For the product $A \times B$,

(11) $$|A \times B| = |A|\,|B|.$$

Finally

(12) $$|\phi| = 0.$$

(13) Example The Shelmikedmu are an elusive and nomadic tribe whose members are unusually heterogeneous in respect of hair and eye colour, and skull shape. A persistent anthropologist establishes the following facts:

(i) 75% have dark hair, the rest have fair hair;
(ii) 80% have brown eyes, the rest have blue eyes;
(iii) no narrow-headed person has fair hair and blue eyes;
(iv) the proportion of blue-eyed broad-headed tribespeople is the same as the proportion of blue-eyed narrow-headed tribespeople;
(v) those who are blue-eyed and broad-headed are fair-haired or dark-haired in equal proportion;
(vi) half the tribe is dark-haired and broad-headed;
(vii) the proportion who are brown-eyed, fair-haired and broad-headed is equal to the proportion who are brown eyed, dark-haired and narrow-headed.

The anthropologist also finds n, the proportion of the tribe who are narrow-headed, but unfortunately this information is lost in a clash with a crocodile on the difficult journey home. Is another research grant and field trip required to find n? Fortunately not, if the anthropologist uses set theory. Let

B be the set of those with blue eyes
C the set of those with narrow heads
D the set of those with dark hair.

Then the division of the tribe into its heterogeneous sets can be represented by Figure 1.3. This type of representation of sets and their relationships is known as a *Venn diagram*. The proportion of the population in each set is denoted by the lower case letter in each compartment, so

$$a = |B^c \cap C^c \cap D^c|/|\Omega|,$$
$$b = |B \cap C^c \cap D^c|/|\Omega|,$$

and so on. The required proportion having narrow heads is

$$n = |C|/|\Omega| = e + f + g + h$$

and of course $a + b + c + d + e + f + g + h = 1$. The information in (i)–(vii), which survived the crocodile, yields the following relationships:

(i) $c + d + e + f = 0.75$
(ii) $a + d + e + h = 0.8$
(iii) $g = 0$
(iv) $f + g = b + c$
(v) $b = c$
(vi) $c + d = 0.5$
(vii) $a = e$

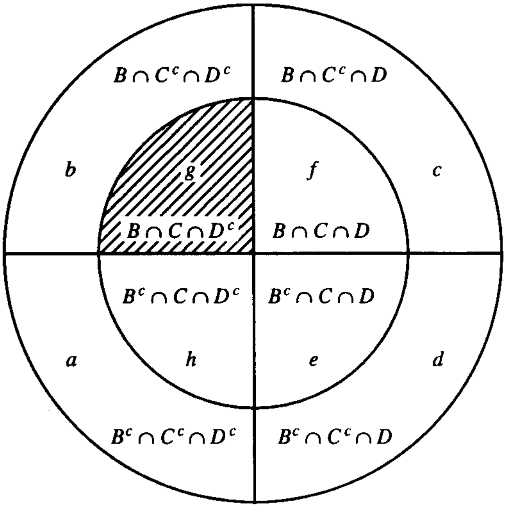

Figure 1.3 Here the interior of the large circle represents the entire tribe, and the interior of the small circle represents those with narrow heads. The part to the right of the vertical line represents those with dark hair, and that part above the horizontal line represents those with blue eyes. Thus the shaded quadrant represents those with blue eyes, narrow heads and fair hair, and as it happens this set is empty by (iii). That is to say $B \cap C \cap D^c = \phi$, and so $g = 0$.

The anthropologist (who has a pretty competent knowledge of algebra) solves this set of equations to find that

$$n = e + f + g + h = e + f + h = 0.15 + 0.1 + 0.05 = 0.3$$

Thus three tenths of the tribe are narrow-headed. ●

This section concludes with a technical note (which you may omit on a first reading). We have noted that events are subsets of Ω. A natural question is, which subsets of Ω are entitled to be called events?

It seems obvious that if A and B are events then all of $A \cup B$, A^c, $A \cap B$, and so on should be entitled to be events also. This is a bit vague, to be precise, we say that a subset A of Ω can be an event if it belongs to a collection $\mathcal{F}$ of subsets of Ω obeying the following three rules:

(14) $$\Omega \in \mathcal{F};$$

(15) $$\text{if } A \in \mathcal{F} \text{ then } A^c \in \mathcal{F};$$

(16) $$\text{if } A_j \in \mathcal{F} \text{ for } j \geqslant 1, \text{ then } \bigcup_{j=1}^{\infty} A_j \in \mathcal{F}.$$

The collection $\mathcal{F}$ is called an event space, or a σ-field.

Notice that using (1)–(6) shows that if A and B are in $\mathcal{F}$ then so are $A \backslash B$, $A \triangle B$ and $A \cap B$.

(17) Example (7) revisited It is easy for you to check that $\{\phi, A, B, \Omega\}$ is an event space, and $\{\phi, A \cup C, B \backslash C, \Omega\}$ is an event space. However, $\{\phi, A, \Omega\}$ and $\{\phi, A, B, D, \Omega\}$ are *not* event spaces. ●

In general, if Ω is finite it is quite usual to take $\mathcal{F}$ to be the collection of all subsets of Ω, which is clearly an event space. If Ω is infinite then this collection is sometimes too big to be useful, and some smaller collection of subsets is required.

1.3 Symmetry and probability

Following the program sketched out in Section 1.1, our next task is to assign a probability $\mathbf{P}(A)$ to any event A in the event space $\mathcal{F}$, and then to define the relationships between these probabilities. In this we are guided by experience and intuition.

Since childhood, for most of us, our intuition has been strongly guided by a special class of experiments which are possessed of an intrinsic symmetry. When rolling a well-made die, we feel intuitively that the symmetries of the die (under reflections and rotations) should find empirical expression in each face being equally likely to end uppermost. Using the notation of Example 1.2.7, and writing $\mathbf{P}(\{\omega_j\})$ for the probability that the jth face is uppermost, this entails

$$\mathbf{P}(\{\omega_i\}) = \mathbf{P}(\{\omega_j\}),$$

$$\mathbf{P}(\{\omega_1, \omega_2\}) = \mathbf{P}(\{\omega_3, \omega_4\}),$$

and so on.

Recalling the convention that probabilities lie between zero and one, it is tempting and natural, given an event A, to set

(1) $$\mathbf{P}(A) = \frac{|A|}{|\Omega|},$$

so that for any $A \subseteq \Omega$, we do indeed have

(2) $$0 \leq \mathbf{P}(A) \leq 1.$$

(3) **Example (1.2.7) revisited** With the probability of any event now defined by (1), it is elementary to find the probability of any of the events that may occur when we roll a die. The probability that it shows an even number is

$$\mathbf{P}(A) = \mathbf{P}(\{\omega_2, \omega_4, \omega_6\}) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

Likewise, and equally trivially, we find that

$$\mathbf{P}(\text{odd}) = \mathbf{P}(B) = \frac{1}{2}$$

$$\mathbf{P}(\text{prime}) = \mathbf{P}(C) = \mathbf{P}(\{\omega_2, \omega_3, \omega_5\}) = \frac{1}{2}$$

$$\mathbf{P}(\text{perfect}) = \mathbf{P}(D) = \mathbf{P}(\{\omega_6\}) = \frac{1}{6}.$$

These values of the probabilities are not inconsistent with our ideas about how the symmetries of this die should express themselves when it is rolled. ●

Recalling the properties of the size $|A|$ of an event A, listed in (1.2.8), (1.2.9), (1.2.10) and (1.2.12), we see that the function $\mathbf{P}(.)$ defined by (1) has the following properties.

For any events A and B

(4) $$\mathbf{P}(A \cup B) + \mathbf{P}(A \cap B) = \mathbf{P}(A) + \mathbf{P}(B).$$

In particular, if A and B are disjoint then

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

More generally, if $A_1, A_2, \ldots, A_n$ form a collection of disjoint events (which is to say that $A_i \cap A_j = \phi$ when $i \neq j$), then

(5) $$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbf{P}(A_i).$$

This property (5) is known as *finite additivity*.

If $A \subseteq B$, then

(6) $$\mathbf{P}(A) \leq \mathbf{P}(B)$$

and finally

(7) $$\mathbf{P}(\phi) = 0.$$

Once again, these statements are quite consistent with our intuition about likelihoods.

Historically the theory of probability has its roots firmly based in just this kind of speculation about games of chance employing cards, dice and lotteries.

(8) **Example** Three dice are rolled and the numbers on the upper faces are added together. The outcomes 9 and 10 can each be obtained in six distinct ways, thus:

$$10 = 1 + 3 + 6 = 1 + 4 + 5 = 2 + 2 + 6 = 2 + 3 + 5 = 2 + 4 + 4 = 3 + 3 + 4$$

$$9 = 1 + 2 + 6 = 1 + 3 + 5 = 1 + 4 + 4 = 2 + 2 + 5 = 2 + 3 + 4 = 3 + 3 + 3.$$

Some time before 1642, Galileo was asked to explain why, despite this, the outcome 10 is more likely than the outcome 9, as shown by repeated experiment. He observed that the sample space Ω has in fact $6^3 = 216$ outcomes, being all possible triples of numbers from 1 to 6. Of these, 27 sum to 10, and 25 sum to 9, so $\mathbf{P}(10) = \frac{27}{216}$ and $\mathbf{P}(9) = \frac{25}{216}$.

This provides an explanation for the preponderance of 10 over 9. ●

It is just this kind of agreement between theory and experiment that justifies our rather arbitrary adoption of rules (4)–(7) above. We shall see many more examples of this.

1.4 Properties of probability

Most experiments do not possess strong symmetry, nor is it often plausible on other grounds that all outcomes should be equally likely. We are nevertheless still faced with the problem of defining a probability function $\mathbf{P}(.): \mathcal{F} \to [0, 1]$ so that $\mathbf{P}(A)$ tells us how likely the event A is to occur.

Here our intuition doesn't tell us much. However, it seems unthinkable that our function, when we produce it, should not be relevant when we roll a die. It is thus natural and convenient, and as you will discover very rewarding, to make the following definition, guided by the discussion of Section 1.3.

Definition The function $\mathbf{P}(.): \mathcal{F} \to [0, 1]$ is a *probability function* if

(1) $$\mathbf{P}(A) \geqslant 0 \text{ for all } A \in \mathcal{F},$$

(2) $$\mathbf{P}(\Omega) = 1$$

and

(3) $$\mathbf{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbf{P}(A_j)$$

whenever $A_1, A_2, \ldots$ are disjoint events (which is to say that $A_i \cap A_j = \phi$ whenever $i \neq j$). ▲

In passing, we note that (3) is known as the property of *countable additivity*. Obviously it implies *finite additivity* so that, in particular, if $A \cap B = \phi$, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.

From these three rules we can derive many important and useful relationships, for example:

(4) $$\mathbf{P}(\phi) = 0,$$

$$\mathbf{P}(A^c) = 1 - \mathbf{P}(A), \tag{5}$$

$$\mathbf{P}(A \backslash B) = \mathbf{P}(A) - \mathbf{P}(A \cap B), \tag{6}$$

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B), \tag{7}$$

$$\begin{aligned}\mathbf{P}\left(\bigcup_1^n A_i\right) = \sum_1^n \mathbf{P}(A_i) - \sum_{i<j} \mathbf{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbf{P}(A_i \cap A_j \cap A_k) + \dots \\ + (-)^{n+1}\mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n).\end{aligned} \tag{8}$$

The following examples begin to demonstrate the importance and utility of (1)–(8).

(9) Example Let us prove (4), (5) and (7) above. First, by (2) and (3), for any $A \in \mathcal{F}$,

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup A^c) = \mathbf{P}(A) + \mathbf{P}(A^c)$$

which proves (5). Now setting $A = \Omega$ establishes (4). Finally, using (3) repeatedly we obtain

$$\mathbf{P}(B) = \mathbf{P}(B \cap A) + \mathbf{P}(B \cap A^c), \tag{10}$$

and

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \cup (B \cap A^c)) = \mathbf{P}(A) + \mathbf{P}(B \cap A^c) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(B \cap A) \quad \text{by (1)},$$

which proves (7). ●

You should now prove (6) as an elementary exercise; the proof of (8) is part of Problem 12.

(11) Example: Inequalities for P(.) (i) If $A \subseteq B$ then $B \cap A = A$, so from (10)

$$\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \cap A^c) \geqslant \mathbf{P}(A) \qquad \text{by (1).}$$

(ii) For any A, B we have *Boole's inequalities*

$$\begin{aligned} \mathbf{P}(A) + \mathbf{P}(B) &\geqslant \mathbf{P}(A \cup B) && \text{by (7)} \\ &\geqslant \max\{\mathbf{P}(A), \mathbf{P}(B)\} && \text{by part (i)} \\ &\geqslant \mathbf{P}(A \cap B) && \text{by part (i) again} \\ &\geqslant \mathbf{P}(A) + \mathbf{P}(B) - 1 && \text{by (7) again.} \end{aligned} \tag{12}$$

●

(13) Example: Lottery An urn contains 1000 lottery tickets numbered from 1 to 1000. One is selected at random. A fairground performer offers to pay \$3 to anyone who has already paid him \$2, if the number on the ticket is divisible by 2, 3 or 5. Would you pay him your \$2 before the draw? (If the ticket number is not divisible by 2, 3 or 5 you lose your \$2.)

Solution Let D_k be the event that the number drawn is divisible by k. Then

$$\mathbf{P}(D_2) = \frac{500}{1000} = \frac{1}{2}$$

and so on. Also

$$\mathbf{P}(D_2 \cap D_3) = \mathbf{P}(D_6) = \frac{166}{1000}$$

and so forth. Using (8) with $n = 3$, and making several similar calculations we have

$$\begin{aligned}\mathbf{P}(D_2 \cup D_3 \cup D_5) &= \mathbf{P}(D_2) + \mathbf{P}(D_3) + \mathbf{P}(D_5) + \mathbf{P}(D_2 \cap D_3 \cap D_5)\\ &\quad -\mathbf{P}(D_2 \cap D_3) - \mathbf{P}(D_3 \cap D_5) - \mathbf{P}(D_2 \cap D_5)\\ &= 10^{-3}(500 + 333 + 200 + 33 - 166 - 66 - 100) = \frac{367}{500}.\end{aligned}$$

The odds on winning are thus better than 2:1, and you should accept his very generous offer. ●

1.5 Sequences of events

This section is important, but may be omitted at a first reading.

Very often we will be confronted by an infinite sequence of events $(A_n;\ n \geqslant 1)$ such that $A = \lim_{n\to\infty} A_n$ exists. In particular, if $A_n \subseteq A_{n+1}$ for all n then

(1) $$\lim_{n\to\infty} A_n = \bigcup_{j=1}^{\infty} A_j = A,$$

and A is an event by (1.2.16). It is of interest, and also often useful, to know $\mathbf{P}(A)$. The following theorem is therefore important as well as attractive.

(2) Theorem If $A_n \subseteq A_{n+1} \in \mathcal{F}$ for all $n \geqslant 1$, and $A = \lim_{n\to\infty} A_n$, then

$$\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

Proof Since $A_n \subseteq A_{n+1}$, we have $(A_{j+1}\backslash A_j) \cap (A_{k+1}\backslash A_k) = \phi$, for $k \neq j$. Also, setting $A_0 = \phi$

$$\bigcup_{1}^{n}(A_j\backslash A_{j-1}) = A_n.$$

Furthermore,

(3) $$\mathbf{P}(A_{n+1}\backslash A_n) = \mathbf{P}(A_{n+1}) - \mathbf{P}(A_n).$$

Hence, since $A_0 = \phi$

$$\begin{aligned}\mathbf{P}(A) = \mathbf{P}(\lim_{n\to\infty} A_n) &= \mathbf{P}(\bigcup_{j=0}^{\infty} A_{j+1}\backslash A_j)\\ &= \sum_{j=0}^{\infty} \mathbf{P}(A_{j+1}\backslash A_j) && \text{by (1.4.3)}\\ &= \lim_{n\to\infty} \mathbf{P}(A_n) && \text{by (3).}\end{aligned}$$ ■

From this result it is a simple matter to deduce that if $A_n \supseteq A_{n+1}$ for all n then

(4) $$\lim_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(A).$$

With a bit more work one can show more generally that if $\lim_{n\to\infty} A_n = A$, then (4) is still true. Because of this, the function $\mathbf{P}(.)$ is said to be a *continuous set function*.

1.6 Remarks

Simple problems in probability typically require the calculation of the probability $\mathbf{P}(E)$ of some event E, or at least the calculation of bounds for $\mathbf{P}(E)$. The underlying experiment may be implicit or explicit; in any case the first step is always to choose the sample space. (Naturally, we choose the one which makes finding $\mathbf{P}(E)$ easiest.)

Then you may set about finding $\mathbf{P}(E)$ by using the rules and results of Section 1.4, and the usual methods for manipulating sets. Useful aids at this simple level include Venn diagrams, and identities such as de Morgan's laws, namely:

(1) $$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$$

and

(2) $$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c.$$

These are readily established directly, or by induction as follows. First, draw a Venn diagram to see that $(A \cup B)^c = A^c \cap B^c$.

Now if $\bigcup_{j=1}^{n} A_j = B_n$ then

$$\left(\bigcup_{j=1}^{n+1} A_j\right)^c = (A_{n+1} \cup B_n)^c = A_{n+1}^c \cap B_n^c.$$

Hence (1) follows by induction on n. The result (2) can also be proved directly, or by induction, or by using (1). You can do this, and you should do it now.

We end this section with a note for the more demanding reader. It has been claimed above that probability theory is of wide applicability, yet most of the examples in this chapter deal with the behaviour of coins, dice, lotteries and urns. For most of us, weeks or even months can pass without any involvement with dice or urns. (The author has never even seen an urn, let alone removed a red ball from one.) The point is that such simple problems present their probabilistic features to the reader unencumbered by strained assumptions about implausible models of reality. The penalty for simplicity is that popular problems may become hackneyed through overuse in textbooks. We take this risk, but reassure the reader that more realistic problems feature largely in later chapters. In addition, when considering some die or urn, students may be pleased to know that they are treading in the footsteps of many very eminent mathematicians as they perform these calculations. Euler or Laplace may have pondered over exactly the same difficulty as you, though not for so long perhaps.

WORKED EXAMPLES AND EXERCISES

1.7 Example: Dice

You roll two dice. What is the probability of the events:

(a) they show the same?
(b) their sum is seven or eleven?
(c) they have no common factor greater than unity?

Solution First we must choose the sample space. A natural representation is as ordered pairs of numbers (i, j), where each number refers to the face shown by one of the dice. We require the dice to be distinguishable (one red and one green say), so that $1 \leqslant i \leqslant 6$ and $1 \leqslant j \leqslant 6$. Because of the symmetry of a perfect die, we assume that these 36 outcomes are equally likely.

(a) Of these 36 outcomes, just 6 are of the form (i, i), so using (1.3.1) the required probability is

$$\frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}.$$

(b) There are 6 outcomes of the form $(i, 7 - i)$, whose sum is 7, so the probability that the sum is 7 is $\frac{6}{36} = \frac{1}{6}$.

There are 2 outcomes whose sum is 11, namely $(5, 6)$ and $(6, 5)$, so the probability that the sum is 11 is $\frac{2}{36} = \frac{1}{18}$.

Hence, using (1.4.3), the required probability is $\frac{1}{6} + \frac{1}{18} = \frac{2}{9}$.

(c) It is routine to list the outcomes which do have a common factor greater than unity. They are thirteen in number, namely:

$$\{(i, i); i \geqslant 2\}, (2, 4), (4, 2), (2, 6), (6, 2), (3, 6), (6, 3), (4, 6), (6, 4).$$

This is the complementary event, so by (1.4.5), the required probability is

$$1 - \frac{13}{36} = \frac{23}{36}.$$

Doing it this way gives a slightly quicker enumeration than the direct approach.

(1) Exercise What is the probability that the sum of the numbers is 2, 3, or 12?

(2) Exercise What is the probability that:

(a) the sum is odd?
(b) the difference is odd?
(c) the product is odd?

(3) Exercise What is the probability that one number divides the other?

(4) Exercise What is the probability that the first die shows a smaller number than the second?

(5) Exercise What is the probability that different numbers are shown and the smaller of the two numbers is r, $1 \leqslant r \leqslant 6$?

Remark It was important to distinguish the two dice. Had we not done so, the sample space would have been $\{(i, j); 1 \leqslant i \leqslant j \leqslant 6\}$, and these 21 outcomes are not equally likely, either intuitively or empirically. Note that the dice need not be different colours in fact, it is enough for us to be able to suppose that they are.

1.8 Example: Urn

An urn contains n heliotrope and n tangerine balls. Two balls are removed from the urn together, at random.

(a) What is the sample space?
(b) What is the probability of drawing two balls of different colours?
(c) Find the probability p_n that the balls are the same colour, and evaluate $\lim_{n\to\infty} p_n$.

Solution (a) As balls of the same colour are otherwise indistinguishable, and they are drawn together, a natural sample space is $\Omega = \{HH, HT, TT\}$.

(b) The outcomes in Ω exhibit no symmetry. Taking our cue from the previous example, we choose to distinguish the balls by numbering them from 1 to $2n$, and also suppose that they are drawn successively. Then the sample space is the collection of ordered pairs of the form (i, j) where $1 \le i, j \le 2n$, and $i \ne j$, because we cannot pick the same ball twice. These $2n(2n-1)$ outcomes are equally likely, by symmetry. In n^2 of them we draw heliotrope followed by tangerine, and in n^2 we draw tangerine followed by heliotrope. Hence

$$\mathbf{P}(HT) = \frac{n^2 + n^2}{2n(2n-1)} = \frac{n}{2n-1}.$$

(c) By (1.4.5)

$$p_n = 1 - \mathbf{P}(HT) = \frac{n-1}{2n-1}$$
$$\to \frac{1}{2} \quad \text{as } n \to \infty.$$

(1) **Exercise** Find $\mathbf{P}(HH)$ when the sample space is taken to be all unordered pairs of distinguishable balls.

(2) **Exercise** What is the probability that (a) the first ball is tangerine? (b) the second ball is tangerine?

(3) **Exercise** Half the balls are removed and placed in a box. One of those remaining in the urn is removed. What is the probability that it is tangerine?

(4) **Exercise** A fair die with n sides is rolled. If the rth face is shown, r balls are removed from the urn and placed in a bag. What is the probability that a ball removed at random from the bag is tangerine?

1.9 Example: Cups and saucers

A teaset has four cups and saucers with two cups and saucers in each of two different colours. If the cups are placed at random on the saucers, what is the probability that no cup is on a saucer of the same colour?

***Solution* I** Call the colours azure and blue; let A be the event that an azure cup is on an azure saucer, and B the event that a blue cup is on a blue saucer. Since there are only two places for blue cups not to be on blue saucers, we see that A occurs if and only if B occurs, so $A = B$ and

$$\mathbf{P}((A \cup B)^c) = \mathbf{P}(A^c) = 1 - \mathbf{P}(A).$$

Now $\mathbf{P}(A) = \mathbf{P}(A_1 \cup A_2)$, where A_1 and A_2 denote the events that the first and second azure cups respectively are on azure saucers. There are 24 equally probable ways of putting the four cups on the four saucers. In 12 of them A_1 occurs; in 12 of them A_2 occurs; and in 4 of them $A_1 \cap A_2$ occurs, by enumeration.

Hence, by (1.4.7),

$$\mathbf{P}(A) = \frac{12}{24} + \frac{12}{24} - \frac{4}{24} = \frac{5}{6},$$

and the required probability is $\frac{1}{6}$.

***Solution* II** Alternatively, instead of considering all the ways of placing cups on saucers, we may consider only the distinct ways of arranging the cups by colour with the saucers fixed. There are only six of these, namely:

$$aabb;\ abba;\ abab;\ baab;\ baba;\ bbaa;$$

and by symmetry they are equally likely. By inspection, in only one of these arrangements is no cup on a saucer of the same colour, so the required probability is $\frac{1}{6}$.

Remark In this example, considering a *smaller* sample space makes the problem easier. This is in contrast to our solutions to Examples 1.7 and 1.8 where we used larger sample spaces to simplify things.

(1) **Exercise** What is the probability that exactly
(a) one cup is on a saucer of the same colour?
(b) two cups are on saucers of the same colour?

(2) **Exercise** What is the probability that no cup is on a saucer of the same colour if the set comprises four cups and saucers in four distinct colours?

1.10 Example: Sixes

Three players, Achilles, Briseis and Chryseis, take it in turns to roll a die in the order ABC, ABC, A . . . Each player drops out of the game immediately upon throwing a six.

(a) What is the sample space for this experiment?
(b) Suppose the game stops when two players have rolled a six. What is the sample space for this experiment?
(c) What is the probability that Achilles is the second player to roll a six?
(d) Let D_n be the event that the third player to roll a six does so on the nth roll. Describe the event E given by

$$E = \left(\bigcup_{n=1}^{\infty} D_n\right)^c.$$

Solution (a) Let U be the collection of all sequences $x_1, \ldots, x_n$ for all $n \geq 1$, such that

$$x_j \in \{1, 2, 3, 4, 5\} \quad \text{for } 1 \leqslant j < n,$$
$$x_n = 6.$$

Then each player's rolls generate such a sequence, and the sample space consists of all selections of the triple (u_1, u_2, u_3) where $u_i \in U$ for $1 \leqslant i \leqslant 3$. We may denote this by $U \times U \times U$, or even U^3 if we wish.

(b) Let V be the collection of all sequences $x_1, x_2, \ldots, x_n$, for $n \geqslant 1$, such that

$$x_j \in \{1, 2, 3, 4, 5\} \quad \text{for } 1 \leqslant j \leqslant n.$$

Then the sample space consists of two selections u_1, u_2 from U, corresponding to the players who roll sixes, and one selection v from V corresponding to the player who does not. The length of v equals the longer of u_1 and u_2 if this turn in the round comes before the second player to get a six, or it is one less than the longer of u_1 and u_2 if this turn in the round is later than the second player to get a six.

(c) Despite the answers to (a) and (b) we use a different sample space to answer this question. Suppose that the player who is first to roll a six continues to roll the die when the turn comes round, these rolls being ignored by the others. This does not affect the respective chances of the other two to be the next player (of these two) to roll a six. We therefore let Ω be the sample space consisting of all sequences of length $3r + 1$, for $r \geqslant 0$, using the integers 1, 2, 3, 4, 5 or 6. This represents $3r + 1$ rolls of the die, and by the assumed symmetry the 6^{3r+1} possible outcomes are all equally likely for each r.

Suppose Achilles is the second player to roll a six on the $3r + 1$th roll. Then his $r + 1$ rolls include no six except his last roll; this can occur in 5^r ways. If Briseis was first to roll a six, then her r rolls include at least one six; this may be accomplished in $6^r - 5^r$ ways. In this case Chryseis rolled no six in r attempts; this can be done in 5^r ways. Hence Achilles is second to Briseis in $5^r.5^r.(6^r - 5^r)$ outcomes. Likewise he is second to Chryseis in $5^r.5^r.(6^r - 5^r)$ outcomes. Hence the probability that he is second to roll a six on the $3r + 1$th roll is

$$p_r = \frac{2(6^r - 5^r)5^{2r}}{6^{3r+1}}, \quad \text{for } r \geqslant 1.$$

By (1.4.3) therefore, the total probability that Achilles is second to roll a six is the sum of these, namely

$$\sum_{r=1}^{\infty} p_r = \frac{300}{1001}.$$

(d) The event $\bigcup_{n=1}^{\infty} D_n$ is the event that the game stops at the nth roll for some $n \geqslant 1$. Therefore E is the event that they never stop.

(1) **Exercise** For each of the three players, find the probability that he or she is the first to roll a six.

(2) **Exercise** Show that $\mathbf{P}(E) = 0$.

(3) **Exercise** Find the probability that the Achilles rolls a six before Briseis rolls a six. [**Hint**: use a *smaller* sample space.]

(4) **Exercise** Show that the probability that Achilles is last to throw a six is $\frac{305}{1001}$. Are you surprised that he is more likely to be last than to be second?

Remark The interesting thing about the solution to (c) is that the sample space Ω includes outcomes that are not in the original experiment, whose sample space is described in (a). The point is that the event in question has the same probability in the original experiment and in the modified experiment, but the required probability is obtained rather more easily in the second case because the outcomes are equally likely. This idea of augmenting the sample space was first used by Pascal and Fermat in the seventeenth century.

In fact we shall find easier methods for evaluating this probability in Chapter 2, using new concepts.

1.11 Example: Family planning

A woman planning her family considers the following schemes, on the assumption that boys and girls are equally likely at each delivery:

(a) have three children;
(b) bear children until the first girl is born or until three are born, whichever is sooner, and then stop;
(c) bear children until there is one of each sex or until there are three, whichever is sooner and then stop.

Let B_i denote the event that i boys are born, and let C denote the event that more girls are born than boys. Find $\mathbf{P}(B_1)$ and $\mathbf{P}(C)$ in each of the cases (a) and (b).

Solution (a) If we do not consider order there are four distinct possible families: BBB; GGG; GGB; BBG, but these are not equally likely. With order included, there are eight possible families in this larger sample space:

$$\{BBB; BBG; BGB; GBB; GGB; GBG; BGG; GGG\} = \Omega \tag{1}$$

and by symmetry they are equally likely. Now, by (1.3.1), $\mathbf{P}(B_1) = \frac{3}{8}$ and $\mathbf{P}(C) = \frac{1}{2}$. The fact that $\mathbf{P}(C) = \frac{1}{2}$ is also clear by symmetry.

Now consider (b). There are four possible families: $F_1 = G$, $F_2 = BG$, $F_3 = BBG$, and $F_4 = BBB$.

Once again these outcomes are not equally likely, but as we have now done several times we can use a different sample space. One way is to use the sample space in (1), remembering that if we do this then some of the later births are fictitious. The advantage is that outcomes are equally likely by symmetry. With this choice F_2 corresponds to $\{BGG \cup BGB\}$ and so $\mathbf{P}(B_1) = \mathbf{P}(F_2) = \frac{1}{4}$. Likewise $F_1 = \{GGG \cup GGB \cup GBG \cup GBB\}$ and so $\mathbf{P}(C) = \frac{1}{2}$.

(2) **Exercise** Find $\mathbf{P}(B_1)$ and $\mathbf{P}(C)$ in case (c).

(3) **Exercise** Find $\mathbf{P}(B_2)$ and $\mathbf{P}(B_3)$ in all three cases.

(4) **Exercise** Let E be the event that the completed family contains equal numbers of boys and girls. Find $\mathbf{P}(E)$ in all three cases.

1.12 Example: Craps

You roll two fair dice. If the sum of the numbers shown is 7 or 11 you win, if it is 2, 3 or 12 you lose. If it is any other number j, you continue to roll two dice until the sum is j or 7, whichever is sooner. If it is 7 you lose, if it is j you win. What is the probability p that you win?

Solution Suppose that you roll the dice n times. That experiment is equivalent to rolling $2n$ fair dice, with the sample space Ω_{2n} being all possible sequences of length $2n$, of the numbers 1, 2, 3, 4, 5, 6, for any $n \geqslant 1$. By symmetry these 6^{2n} outcomes are equally likely, and whether you win or you lose at or before the nth roll of the pair of dice is determined by looking at the sum of successive pairs of numbers in these outcomes.

The sample space for the roll of a pair of dice (Ω_2), has 36 equally likely outcomes. Let n_j denote the number of outcomes in which the sum of the numbers shown is j, $2 \leqslant j \leqslant 12$. Now let A_k be the event that you win by rolling a pair with sum k, and consider the eleven distinct cases:

(a)
$$\mathbf{P}(A_2) = \mathbf{P}(A_3) = \mathbf{P}(A_{12}) = 0,$$

because you always lose with these.

(b) For A_7 to occur you must get 7 on the first roll. Since $n_7 = 6$

$$\mathbf{P}(A_7) = \frac{n_7}{|\Omega_2|} = \frac{6}{36} = \frac{1}{6}.$$

(c) Likewise

$$\mathbf{P}(A_{11}) = \frac{n_{11}}{36} = \frac{2}{36} = \frac{1}{18}, \quad \text{since } n_{11} = 2.$$

(d) For A_4 to occur you must get 4 on the first roll and on the nth roll, for some $n \geqslant 2$, with no 4 or 7 in the intervening $n-2$ rolls. You can do this in $n_4^2(36 - n_4 - n_7)^{n-2}$ ways, and therefore

$$\mathbf{P}(A_4) = \sum_{n=2}^{\infty} \frac{n_4^2(36 - n_4 - n_7)^{n-2}}{6^{2n}} \qquad \text{by (1.4.3)}$$

$$= \frac{n_4^2}{36(n_4 + n_7)} = \frac{1}{36} \qquad \text{because } n_4 = 3.$$

(e) Likewise

$$\mathbf{P}(A_5) = \frac{n_5^2}{36(n_5 + n_7)}$$

$$= \frac{2}{45} \qquad \text{because } n_5 = 4$$

$$= \mathbf{P}(A_9) \qquad \text{because } n_9 = 4.$$

Finally

$$\mathbf{P}(A_6) = \mathbf{P}(A_8)$$

$$= \frac{25}{396} \qquad \text{because } n_6 = n_8 = 5$$

and

$$\mathbf{P}(A_{10}) = \mathbf{P}(A_4) \qquad \text{because } n_{10} = n_4 = 3.$$

Therefore the probability that you win is

$$\mathbf{P}(A_7) + \mathbf{P}(A_{11}) + 2\mathbf{P}(A_4) + 2\mathbf{P}(A_5) + 2\mathbf{P}(A_6) = \frac{1}{6} + \frac{1}{18} + \frac{1}{18} + \frac{4}{45} + \frac{25}{198}$$
$$\simeq 0.493.$$

(1) **Exercise** What is the probability that you win on or before the second roll?

(2) **Exercise** What is the probability that you win on or before the third roll?

(3) **Exercise** What is the probability that you win if, on the first roll,
(a) the first die shows 2?
(b) the first die shows 6?

(4) **Exercise** If you could fix the number to be shown by one die of the two on the first roll, what number would you choose?

1.13 Example: Murphy's law

A fair coin is tossed repeatedly. Let s denote any fixed sequence of heads and tails of length r. Show that with probability one the sequence s will eventually appear in r consecutive tosses of the coin.

[The usual statement of Murphy's law says that anything that can go wrong, will go wrong].

Solution If a fair coin is tossed r times, there are 2^r distinct equally likely outcomes and one of them is s. We consider a fair die with 2^r faces; each face corresponds to one of the 2^r outcomes of tossing the coin r times and one of them is face s. Now roll the die repeatedly.

Let A_k be the event that face s appears for the first time on the kth roll. There are 2^{rk} distinct outcomes of k rolls, and by symmetry they are equally likely. In $(2^r - 1)^{k-1}$ of them A_k occurs, so by (1.3.1)

$$\mathbf{P}(A_k) = \frac{(2^r - 1)^{k-1}}{2^{rk}}.$$

Since $A_k \cap A_j = \phi$ for $k \neq j$ we have by (1.4.3) that

(1) $$\mathbf{P}\left(\bigcup_1^m A_k\right) = \sum_1^m \mathbf{P}(A_k) = 1 - \left(\frac{2^r - 1}{2^r}\right)^m$$

which is the probability that face s appears at all in m rolls.

Now consider n tosses of the coin, and let $m = [\frac{n}{r}]$ (where $[x]$ is the integer part of x). The n tosses can thus be divided up into m sequences of length r with a remainder $n - mr$. Let B_n be the event that none of these m sequences is s, and let C_n be the event that the sequence s does not occur anywhere in the n tosses. Then

$$C_n \subseteq B_n = \left(\bigcup_1^m A_k\right)^c,$$

since rolling the die m times and tossing the coin mr times yield the same sample